EMDataBank.org: unified data resource for CryoEM

Catherine L. Lawson[1],*, Matthew L. Baker[2], Christoph Best[3], Chunxiao Bi[1], Matthew Dougherty[2], Powei Feng[4], Glen van Ginkel[3], Batsal Devkota[1], Ingvar Lagerstedt[3], Steven J. Ludtke[2], Richard H. Newman[3], Tom J. Oldfield[3], Ian Rees[2], Gaurav Sahni[3], Raul Sala[1], Sameer Velankar[3], Joe Warren[4], John D. Westbrook[1], Kim Henrick[3], Gerard J. Kleywegt[3], Helen M. Berman[1] and Wah Chiu[2],*

[1]Department of Chemistry and Chemical Biology and Research Collaboratory for Structural Bioinformatics, Rutgers, The State University of New Jersey, 610 Taylor Road Piscataway, NJ 08854, [2]Department of Biochemistry and Molecular Biology, National Center for Macromolecular Imaging, Verna and Marrs McLean, Baylor College of Medicine, 1 Baylor Plaza, Houston, TX 70030, USA, [3]Protein Data Bank in Europe, EMBL-EBI, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SD, UK and [4]Department of Computer Science, Rice University, 6100 South Main, Houston, TX 77251, USA

ABSTRACT

Cryo-electron microscopy reconstruction methods are uniquely able to reveal structures of many important macromolecules and macromolecular complexes. EMDataBank.org, a joint effort of the Protein Data Bank in Europe (PDBe), the Research Collaboratory for Structural Bioinformatics (RCSB) and the National Center for Macromolecular Imaging (NCMI), is a global 'one-stop shop' resource for deposition and retrieval of cryoEM maps, models and associated metadata. The resource unifies public access to the two major archives containing EM-based structural data: EM Data Bank (EMDB) and Protein Data Bank (PDB), and facilitates use of EM structural data of macromolecules and macromolecular complexes by the wider scientific community.

INTRODUCTION

Cryo-electron microscopy (cryoEM) has become an essential technique in structural biology, bridging the gap between cell biology, X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy (1,2). CryoEM reconstruction methods are being used to determine structures of large macromolecules, macromolecular complexes and cell components involved in many key biological processes including signal transduction, genome replication, transcription, translation, chaperonin-assisted protein folding, viral infection and motility. Three-dimensional (3D) density maps derived from cryoEM experiments reveal overall molecular shape and may be further interpreted through segmentation algorithms, rigid-body fitting of atomic coordinates determined using X-ray crystallography or NMR, and/or *ab initio* model building, depending on map resolution (3–5).

Public access to cryoEM map volumes and their fitted model interpretations permits independent assessment and analysis of structural results and stimulates development of new tools for visualization, fitting and validation. The EM Data Bank (EMDB) is the major repository for 3D map volumes obtained using electron microscopy (6), while the Protein Data Bank (PDB) collects atomic coordinates fitted into EM map volumes (7). The Unified Data Resource for CryoEM (http://EMDataBank.org) was created in order to unify data deposition, processing and retrieval of maps and fitted models. Here, we provide an overview of the EM structural data archives and the unified resource, including their historical context, current content and use, and future prospects.

EM STRUCTURAL DATA ARCHIVES

Maps

The EMDB was established in 2002 by the Macromolecular Structure Database group (now PDBe) at the European Bioinformatics Institute (EBI) in Hinxton, UK (6,8,9), and was initially supported by two European Union-funded projects: Integration of Information about Macromolecular Structure (IIMS) and the 3D Electron Microscopy Network of Excellence (3DEM NoE). A web-based deposition system, EMDEP, was developed to handle data capture (10). EMDEP validates data via an interactive depositor-driven operation, relying on the knowledge and expertise of the experimenters for the complete and accurate description of the structural experiment and its results. The captured metadata (e.g. sample description, specimen preparation, imaging, reconstruction and fitting details) are stored in an XML-style 'header' file, and the deposited map is converted to a common format for redistribution. A database query tool, EMSEARCH, was also designed and implemented to enable web-based searches. By December 2003, the EMDB was an operational public database with 65 maps deposited by major EM laboratories from Europe and the USA. At this time the PDB began to see an increase in EM-related coordinate depositions, in many cases models fitted into maps deposited to EMDB.

Models

The PDB archive was established in 1971 as a public repository for X-ray crystal structures of biological macromolecules (11), and is presently maintained by the global organization world-wide PDB [wwPDB (12)]. The number of structures in the PDB has grown from the initial 7 to over 67 000 entries. Over time, the PDB began to collect coordinates of structures determined by methods other than X-ray crystallography, including NMR spectroscopy, neutron diffraction, fiber diffraction, electron crystallography, electron microscopy and solution scattering. Coordinates for structures determined using EM began to be archived in the 1990s, beginning with models for bacteriorhodopsin (13) and the RecA hexamer (14). Currently, the deposition rate for EM entries in the PDB is increasing more rapidly than for any other experimental method.

EM dictionary development

Two workshops held in 2004 (3DEM NoE workshop at EBI and CryoEM Structure Deposition Workshop at RCSB-PDB, co-sponsored by NCMI) invited the EM community to participate in development of an improved data model for describing cryoEM experiments, and also set in motion efforts to unify deposition and access to EM-derived maps and models. Following the workshops, a revised and expanded EM dictionary handling both map and model metadata was created in a three-way collaboration between PDBe, RCSB and NCMI with broad community input and was presented at the 2005 3DEM Gordon Research Conference in New

Hampshire as well as a 3DEM Developers workshop held in the UK. During this period, Heymann *et al.* (15) also published recommended conventions for exchange of cryoEM data, and a standards task force was created to gather information on the different cryoEM map and image conventions and formats to facilitate conversion.

Unified resource

Following the recommendations of the EM community to create a 'one-stop shop for EM,' the Unified Data Resource for CryoEM (EMDataBank.org) was established in 2007 with funding from the National Institutes of Health/NIGMS as a joint effort of PDBe, RCSB and NCMI. The mission of the resource is to build up a global deposition and retrieval network for cryoEM map, model and associated metadata, as well as a portal for software tools for standardized map format conversion, map, segmentation and model assessment, visualization and data integration. The first goal of this collaboration was achieved in early 2008 when RCSB joined PDBe as a second map deposition and retrieval site. Joint EMDB (map) and PDB (model) deposition systems were developed and put into operation at both PDBe and RCSB in early 2009, and web-based 3D visualization tools have been integrated into EMDB atlas pages. Efforts to improve uniformity and usability of the EM structural data in both EMDB and PDB databases are ongoing; we recently completed remediation of voxel sizes and density statistics stored in all EMDB map files to ensure display at correct physical scale and with reasonable density contour level. Additional services for data harvesting and evaluation are planned.

DEPOSITION AND CONTENT

The EMDB currently holds more than 800 map entries with resolution limits ranging from 80 to 2 Å, while PDB holds more than 300 entries of coordinates fitted into EM map volumes (Figure 1). Map volume and fitted model deposition rates in 2008–09 were ~150 and ~40 per year, respectively, with roughly 40% of all published EM structures being captured in the databases. As the importance of EM-derived structural information continues to increase, it is anticipated that more journals and funding agencies will require deposition to structural databases.
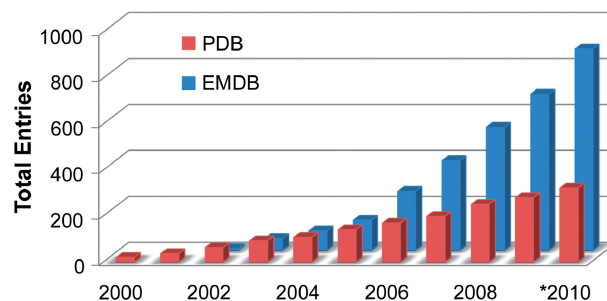

Figure 1. EM entries in EMDB and PDB, cumulative by year (*2010 data is through July).

Each EMDB entry holds a single map plus associated experimental metadata; each PDB entry holds a fitted coordinate model and associated experimental metadata plus primary sequence information for each polymer. The metadata information that is shared by both databases is automatically transferred during the joint deposition process and includes:

- sample description, including components and stoichiometry;
- sample and specimen preparation, including buffer, grid, vitrification, staining;
- imaging experiment, including microscope type, microscope settings, detector;
- specimen selection, image correction, reconstruction method, resolution/resolution method; and
- fitting description, including PDB ids of source coordinates.

Correspondences between maps and associated fitted coordinate models are maintained within both archives. EMDB entries can optionally hold associated masks, structure factors and/or layerline data; PDB entries can also hold structure factors. The underlying dictionaries for the two databases have direct translations and are regularly updated to reflect changes in experimental apparatus and methods.

The map archive includes maps generated by a number of different electron microscopy reconstruction methods (Figure 2). The majority of entries (83%) are single particle reconstructions, which represent ensemble averages of thousands of individual imaged particles, often with additional symmetry averaging. The largest class of single particle specimens represented are the viruses (20% of all holdings), the majority of which are icosahedrally averaged. Virus entries typically represent distinct states of maturation, or complexes with antibodies or receptors. The second largest class of single particle specimens represented are the ribosomes (15%). Ribosome entries define key structural conformations encountered in translating messenger RNA into protein, or elucidate structural variations across diverse species. Other single particle specimens represented include macromolecular machines involved in protein folding, protein degradation, energy metabolism, cell cycle processes,

DNA replication, DNA repair, RNA transcription, RNA splicing and ion channels.

The map archive also holds densities for 2D crystals and for helical arrays, including intracellular filaments and microtubules, flagella and helical crystals. There are also several tomographic maps of unique structures as well as maps that represent 3D averages of aligned tomograms. Diverse specimens currently held of this type include flagellar motors, insect flight muscle tissue, desmosomes and viruses.

EM-derived coordinates are obtained using a variety of modeling methods including manual docking, rigid-body fitting, homology modeling, *de novo* modeling and computational refinement algorithms. EM entries in the PDB are classified either under electron microscopy or electron crystallography as the experimental method. For structures with regular point or helical symmetry, coordinates are given for the asymmetric unit along with a set of transformation matrices to build the biological assembly (16).

A gallery of representative maps including some maps with fitted models is presented in Figure 3.

ACCESS

Access to EM structural data and related services is through the EMDataBank.org web site (Table 1). The EMDB and PDB archives are updated weekly on Wednesdays at 00:00 GMT. EMDB is distributed on ftp mirrors supported by the two EMDataBank.org distribution partners in the UK and the USA, while PDB is distributed on ftp mirrors supported by each of the wwPDB partners. Upon depositor request, EM entries may be held for up to 2 years from the deposition date for map entries in EMDB, and up to 1 year for coordinate models in PDB. However, we strongly encourage depositors to make their data publicly available as soon as possible.

The EMSEARCH web service is maintained and updated weekly at both distribution sites. EMSEARCH enables browsing and searching of EMDB metadata uploaded into a relational database. Simple searches can be performed based on author name, title, entry id, sample name, citation abstract word, aggregation type, resolution and release date range. Search summaries link to a set of atlas pages for each entry, which include 'Summary', 'Visualization', 'Sample', 'Experiment', 'Processing', 'Map Information' and 'Download' pages (Figure 4). In addition, a full-text search system based on the Lucene indexer is nearing completion. All search options can be accessed from the EMDataBank.org search tab.

The 'Visualization' page for each EM map entry provides several viewing options. In addition to a static 2D image provided by the deposition author, it is also possible to launch two different Java-based 3D viewers (Figure 5, left and right top panels). EMViewer provides a simple, single isosurface representation of a map at a predetermined contour level that can be manipulated by mouse drag and click actions. OpenAstexViewer, a molecular graphics program originally developed to display
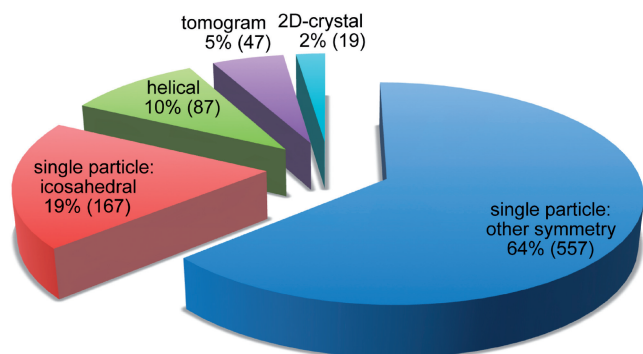

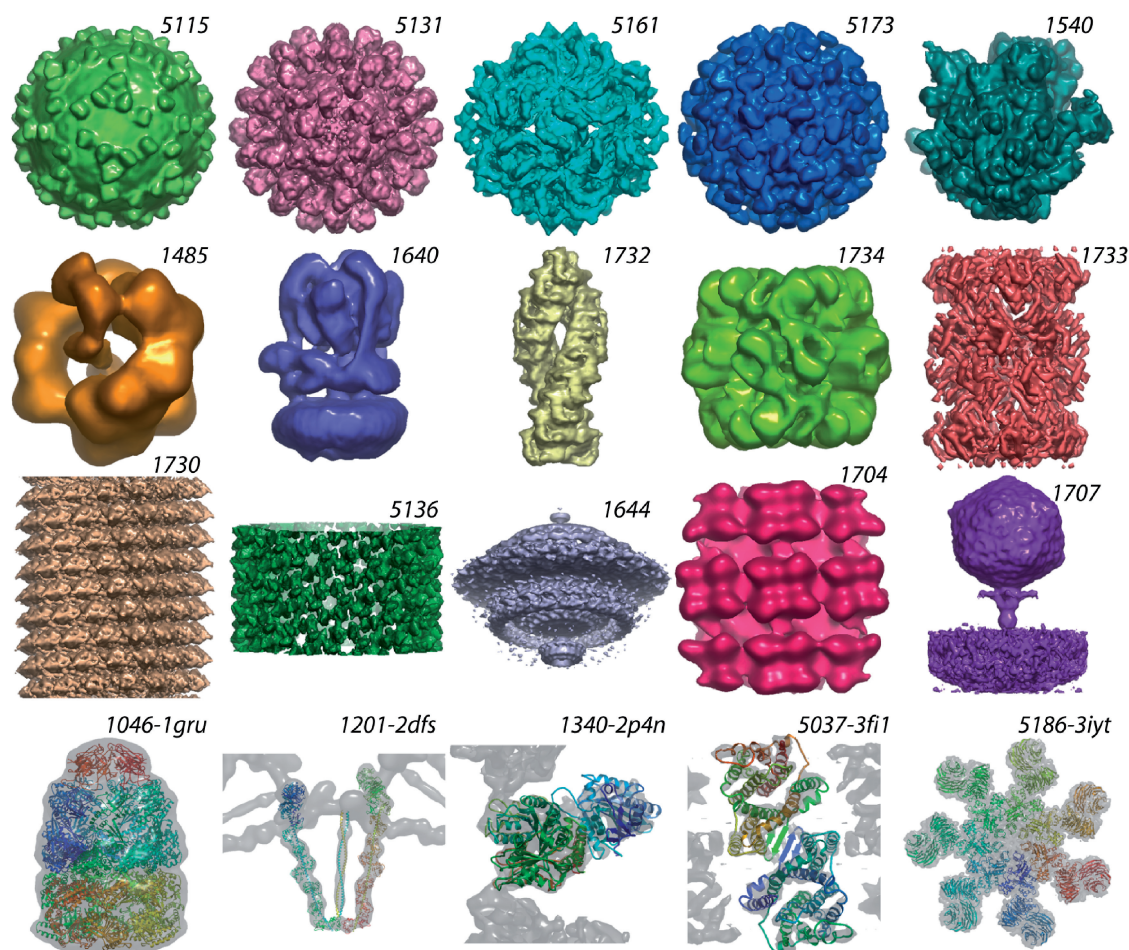**Figure 2.** Current distribution of EM map entry types.

Figure 3. Image gallery of EM structures in EMDB and PDB (labels are database ids with "EMD-" prefixes omitted for brevity; scaling is arbitrary). Icosahedral—*5115*: West Nile virus-antibody complex (53); *5131*: wild-rabbit hemorrhagic disease virus (54); *5161*: Partitivirus (55); *5173*: Hepatitis E particles (56); *other single particle: 1540:* pre-translocational ribosome (57); *1485:* DNA ligase–PCNA–DNA complex (58); *1640:* yeast V-ATPase (59); *1732:* tripeptidyl peptidase II (60); *1734:* rubisco-methyltransferase complex (61); *1733:* 20S proteasome (62); *Helical—1730:* tobacco mosaic virus (63); *5136:* HIV-1 capsid tubular crystal (64); *Tomogram averages—1644:* Borrelia flagellar motor (65); *1704:* Tula Hantavirus glycoprotein spikes (66); *1707:* Podovirus P-SSP7 infecting Prochlorococcus marinus (67); *Map+model pairs—1046-1gru:* GroES–GroEL complex (68); *1201-2dfs:* Myosin V (69); *1340-2p4n:* kinesin bound to a microtubule (70); *5037-3fi1:* ion-coupled transport protein NhaA (71); *5186-3iyt:* apoptosome–procaspase-9 CARD complex (72). All images were created using OpenAstexViewer as implemented on the EMDB atlas pages for these entries.

crystallographic data (17) has been adapted for display of EM maps and associated coordinate models. To improve web download speed and minimize memory requirements, a compact map format is used (BRIX), and larger maps are also down-sampled by a factor of 2–5. Current capabilities include ability to control map contour level, opacity, color, solid versus mesh surface rendering and concurrent display of a PDB coordinate entry. Viewing large maps may require increasing Java Applet Runtime memory allocation.

Maps are distributed in CCP4 format and can be viewed with locally installed software such as UCSF Chimera (18), Pymol (19), VMD (20), Coot (21), enabling further analysis and manipulation with an extensive set of tools. Links for map download are available on the 'Download' and 'Map Information' pages of each entry. Recent distributions of UCSF Chimera enable direct downloads of EMDB maps plus associated PDB models via simple queries (e.g. author name, title) to the EMSEARCH relational database

through the EMDataBank.org beta-web service, a self-contained programing interface based on the SOAP protocol (Figure 5, bottom panel).

EXAMPLES OF USE

A major goal of the EMDataBank.org unified data resource is to archive EM-derived structural information in a way that will enable further research. The impact of availability of structural data for smaller proteins has already been amply demonstrated by the success of the PDB, which is accessed by thousands of individuals around the globe every day. Listed here are a few examples of how archived EM data facilitates subsequent scientific exploration by other investigators.

Building up molecular pictures of large macromolecular assemblies

Atomic coordinates permit 3D mapping of a wide variety of data including sites of mutations leading to disease,

Table 1. Overview of EMDataBank.org resources and services as of September 2010

EMDataBank.org page	Description
/deposit.html	Links to EMDEP deposition servers hosted by RCSB-PDB and PDBe, CryoEM Map+Model Deposition Guide and EMDB Hold/Release Policy
/search.html	Links to EMSEARCH servers hosted by RCSB-PDB and PDBe, Full Text Search (α)
/recententries.html	Table of all recently released and recently submitted EMDB entries, updated weekly
/webservice.html	EMDataBank web service based on the SOAP protocol with application components for use by other applications over the Internet
/otherservices.html	Links to other services including the ftp servers at PDBe (ftp://ftp.ebi.ac.uk/pub/databases/emdb) and RCSB-PDB (ftp://emdb.rutgers.edu)
/allnews.html	EMDataBank.org news items
/faq.html	Frequently asked questions about EM map volume depositions, EM fitted coordinate model depositions and map access
/mapformat.html	EMDB map distribution format description (CCP4 format for 3DEM maps)
/emsoftware.html	List of software available within the EM community for generation, analysis and fitting of EM maps, sortable by program name and type
/emtestdata.html	Links to EM image data sets that have been made publicly available by EM community members to download for testing

Figure 4. Example EMDB atlas page.

modification sites, antibody recognition sites, amino acid sequence variability and electrostatic properties. But in many cases, high-resolution structural models are not available for every component within a cryoEM map at the time it is first interpreted. By preserving map and model information together in a freely accessible database, new structural information can be incorporated as it becomes available. Examples include (i) reinterpretation of ribosome stalk regions in multiple archived maps [EMD-1005 through 1008, 1055, 1056; (22)] and (ii) recent progress by the Rossmann group towards complete fitted coordinate models for bacteriophage T4 in multiple functional states [EMD-1048 and 3H3W; EMD-1086 and 3H3Y (23)].

Structural basis for interpreting data

Structural knowledge can be crucial for interpreting biochemical and biophysical data. For instance, dengue and West Nile virus cryoEM structures (24,25) have led the way to mapping sites of glycosylation (26), amino acid sequence variability (27), neutralizing antibody binding (28), design of vaccines (29) and a structural basis for understanding membrane fusion (30). Fitting of X-ray crystal structures of cadherin into tomographic maps of epidermal desmosomes have yielded insights into cell adhesion [EMD-1051, 1052, 1053, 1374, 1449; (31,32)].

Crystal structure phasing

For low-symmetry complexes such as ribosomes, a low-resolution cryoEM map can provide a valuable complementary source of information for crystallographic phasing. For high-symmetry structures such as icosahedral viruses, application of robust computational averaging and extension algorithms to initial cryoEM-derived phases is often sufficient to complete the structure determination without additional phase
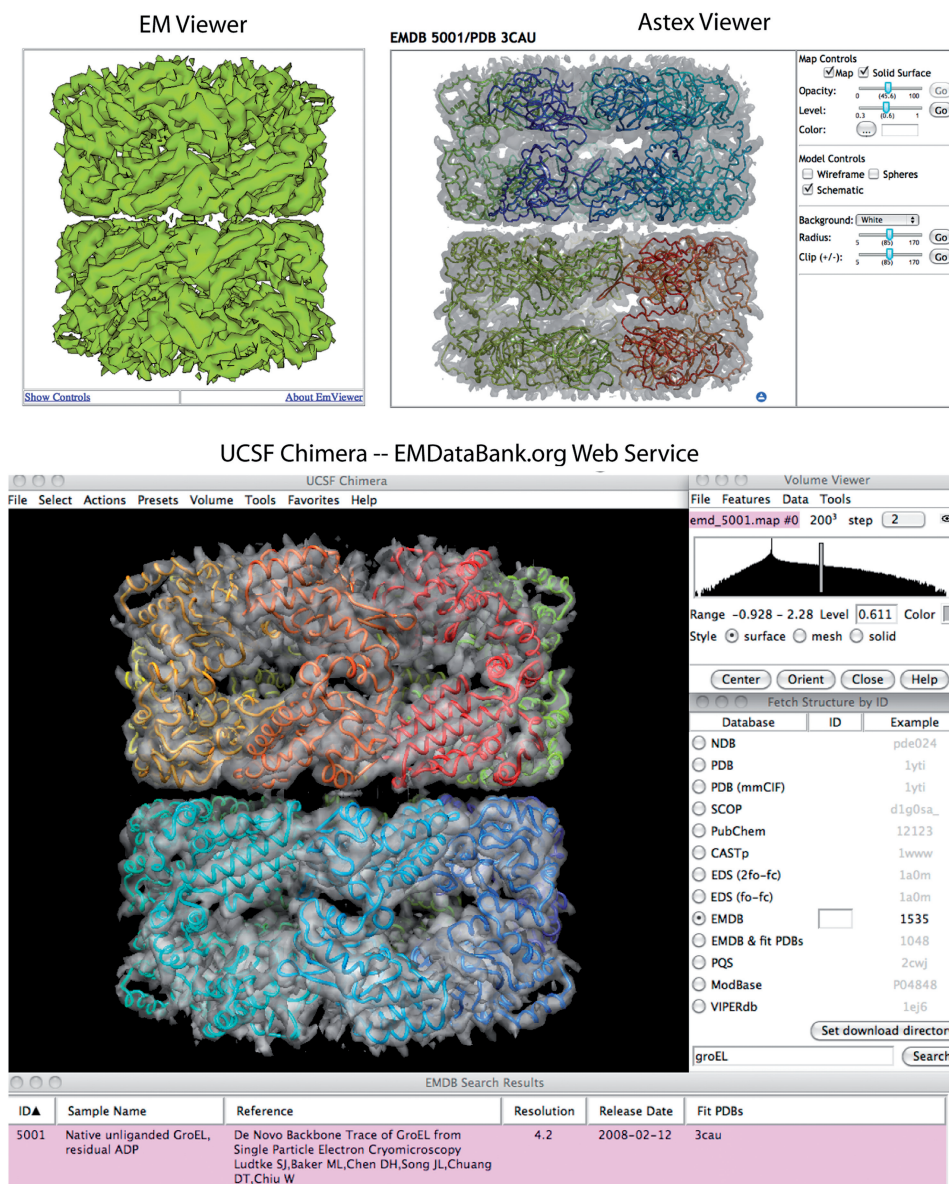
Figure 5. Tools for 3D visualization of EM structural data. EM viewer (top left) and OpenAstexViewer (top right) are Java-based 3D viewers that can be launched from the EMDB atlas 'Visualization' pages. EM viewer employs a compact single contour-level mesh representation for lightweight map viewing; OpenAstexViewer enables exploration of maps and their fitted models with adjustable map contour level. The EMDataBank.org web service implemented in UCSF chimera (bottom panel) expedites search of the EMDB and subsequent map + model download. The example shown is GroEL [EMD-5001and PDB 3CAU; (73)].

information. Recent case studies have employed ribosome maps [EMD-1008 and 1019; (33)] and a dodecameric enzyme map [EMD-1680; (34)]. Routine archiving of EM-derived maps facilitates the use of this method by the crystallography community.

Comparative studies

By capturing coordinates, maps and metadata and making them available through searchable databases, it is possible to compare and select structures for further analysis using a variety of criteria, perform experimental design for new analyses, and design new algorithms to improve the state of the art methodology. For example, comparison of publicly available maps and coordinates led to the

conclusion that the double-stranded DNA tailed phages and herpesvirus have a similar fold in their major capsid proteins though there is little sequence similarity among them [EMD-1101 and PDB entries 1OHG, 1YUE; (35)]. In a second example, a flexible fitting approach applied to 43 maps of bacterial 70S ribosome from EMDB in various functional states revealed global conformational differences between the EM structures involving large-scale ratchet-like deformations (36).

Structures not accessible by other techniques

Many large molecular machines are not amenable to crystallization for high-resolution studies. CryoEM methods have begun to yield structures of membrane proteins,

viruses and chaperonins determined to resolution limits beyond 4.5 Å where *de novo* methods can trace a Cα-backbone. In many cases, side chain densities are also visible. In one recent example, over 80% of side-chain densities in the archaeal chaperonin hexadecamer Mm-cpn chaperonin could be resolved [EMD-5137 and PDB 3LOS; (37)].

Software development

Availability of EM map volumes in the EMDB facilitates development and validation of software for map viewing, analysis, manipulation, coordinate model fitting and validation. Development of algorithms for fitting of EM maps with atomic coordinates is a particularly active area. For low to medium-resolution studies in which the source coordinates are likely to be derived from crystallographic or NMR studies, algorithms being explored include rigid-body fitting (38), normal mode analysis (39), spatial interpolation (40), conformational sampling under low-resolution restraints (37,41), molecular dynamics flexible fitting (42) and simulated annealing approaches (43,44). For medium- to high-resolution studies where *de novo* model building becomes possible, methods are being developed for skeletonization and secondary-structure element detection (45) and incorporation of structure prediction from primary sequence (46). Additional examples of algorithm development facilitated by public availability of EMDB maps include investigations of map denoising (47), B-factor sharpening (48), map resolution determination (49) and automated segmentation (50,51). The UCSF Chimera team has made particularly effective use of the EMDB resource for development and testing of a large, versatile set of tools for manipulating volume data (18,52).

FUTURE PROSPECTS

3D cryoEM reconstruction methods are uniquely able to reveal structural aspects of many important macromolecules and macromolecular complexes, and for this reason the field is in a period of rapid expansion and development. Based on current growth of EM entries and publication of EM structures, the total number of structures of large biological assemblies contributed by EM is anticipated to approach 10 000 by the year 2020.

In the near future, deposition and archiving of EM structural data will be integrated in a common tool that is being developed by the wwPDB partners to handle depositions from all structural biology methods. In addition, as the field matures, validation tools and criteria for assessment of map and fitted coordinate models will play an important role in providing guidance to users of cryoEM derived structural data. To this end, a validation task force is being assembled by the unified data resource partners (NCMI, RCSB and PDBe) to develop recommendations as to how best to assess the quality of both maps and models that have been obtained from cryoEM data. The recommendations will form the basis for a validation suite to be used by EMDB and PDB. The 'CryoEM Modelling Challenge 2010' (http://ncmi.bcm.edu/challenge), in which the modeling community is being asked to apply their modeling tools to a selected set of cryoEM densities at different resolutions, will permit careful comparison of the many modeling methods under development and will yield useful benchmark data for validation suite development.

REFERENCES

1. Frank,J. (2006) *Three-dimensional Electron Microscopy of Macromolecular Assemblies: Visualization of Biological Molecules in their Native State*. Oxford University Press, New York.
2. Glaeser,R.M., Downing,K.H., DeRosier,D., Chiu,W. and Frank,J. (2007) *Electron Crystallography of Biological Macromolecules*. Oxford University Press, Oxford, New York.
3. Chiu,W., Baker,M.L., Jiang,W., Dougherty,M. and Schmid,M.F. (2005) Electron cryomicroscopy of biological machines at subnanometer resolution. *Structure*, **13**, 363–372.
4. Rossmann,M.G., Morais,M.C., Leiman,P.G. and Zhang,W. (2005) Combining x-ray crystallography and electron microscopy. *Structure*, **13**, 355–362.
5. Baker,M.L., Baker,M.R., Hryc,C.F. and Dimaio,F. (2010) Analyses of subnanometer resolution cryoEM density maps. In Jensen,G. (ed.), *Methods in Enzymology*, Vol. 483. Elsevier, London.
6. Tagari,M., Newman,R., Chagoyen,M., Carazo,J.M. and Henrick,K. (2002) New electron microscopy database and deposition system. *Trends Biochem. Sci.*, **27**, 589.
7. Dutta,S., Burkhardt,K., Young,J., Swaminathan,G.J., Matsuura,T., Henrick,K., Nakamura,H. and Berman,H.M. (2009) Data deposition and annotation at the worldwide protein data bank. *Mol. Biotechnol.*, **42**, 1–13.
8. Fuller,S.D. (2003) Depositing electron microscopy maps. *Structure*, **11**, 11–12.
9. Editorial. (2003) *Nat. Struct. Biol.*, **10**, 313.
10. Henrick,K., Newman,R., Tagari,M. and Chagoyen,M. (2003) EMDep: a web-based system for the deposition and validation of high-resolution electron microscopy macromolecular structural information. *J. Struct. Biol.*, **144**, 228–237.
11. Bernstein,F.C., Koetzle,T.F., Williams,G.J.B., Meyer,E.F. Jr, Brice,M.D., Rodgers,J.R., Kennard,O., Shimanouchi,T. and Tasumi,M. (1977) Protein Data Bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.*, **112**, 535–542.
12. Berman,H.M., Henrick,K. and Nakamura,H. (2003) Announcing the worldwide Protein Data Bank. *Nat. Struct. Biol.*, **10**, 980.
13. Henderson,R., Baldwin,J.M., Ceska,T.A., Zemlin,F., Beckmann,E. and Downing,K.H. (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J. Mol. Biol.*, **213**, 899–929.

14. Yu,X. and Egelman,E.H. (1997) The RecA hexamer is a structural homologue of ring helicases. *Nat. Struct. Biol.*, **4**, 101–104.

15. Heymann,J.B., Chagoyen,M. and Belnap,D.M. (2005) Common conventions for interchange and archiving of three-dimensional electron microscopy information in structural biology. *J. Struct. Biol.*, **151**, 196–207.

16. Lawson,C.L., Dutta,S., Westbrook,J.D., Henrick,K. and Berman,H.M. (2008) Representation of viruses in the remediated PDB archive. *Acta Crystallogr. D Biol. Crystallogr.*, **64**, 874–882.

17. Hartshorn,M.J. (2002) AstexViewer: a visualisation aid for structure-based drug design. *J. Comput. Aided Mol. Des.*, **16**, 871–881.

18. Pettersen,E.F., Goddard,T.D., Huang,C.C., Couch,G.S., Greenblatt,D.M., Meng,E.C. and Ferrin,T.E. (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

19. Delano,W.L. (2002) *The Pymol Molecular Graphics System*. Delano Scientific, Palo Alto.

20. Hsin,J., Arkhipov,A., Yin,Y., Stone,J.E. and Schulten,K. (2008) Using VMD: an introductory tutorial. *Curr. Protoc. Bioinformatics*, Chapter 5, Unit 57.

21. Emsley,P. and Cowtan,K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.*, **60**, 2126–2132.

22. Diaconu,M., Kothe,U., Schlunzen,F., Fischer,N., Harms,J.M., Tonevitsky,A.G., Stark,H., Rodnina,M.V. and Wahl,M.C. (2005) Structural basis for the function of the ribosomal L7/12 stalk in factor binding and GTPase activation. *Cell*, **121**, 991–1004.

23. Aksyuk,A.A., Leiman,P.G., Shneider,M.M., Mesyanzhinov,V.V. and Rossmann,M.G. (2009) The structure of gene product 6 of bacteriophage T4, the hinge-pin of the baseplate. *Structure*, **17**, 800–808.

24. Kuhn,R.J., Zhang,W., Rossmann,M.G., Pletnev,S.V., Corver,J., Lenches,E., Jones,C.T., Mukhopadhyay,S., Chipman,P.R., Strauss,E.G. *et al.* (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell*, **108**, 717–725.

25. Mukhopadhyay,S., Kim,B.S., Chipman,P.R., Rossmann,M.G. and Kuhn,R.J. (2003) Structure of West Nile virus. *Science*, **302**, 248.

26. Hanna,S.L., Pierson,T.C., Sanchez,M.D., Ahmed,A.A., Murtadha,M.M. and Doms,R.W. (2005) N-linked glycosylation of west nile virus envelope proteins influences particle assembly and infectivity. *J. Virol.*, **79**, 13262–13274.

27. Modis,Y., Ogata,S., Clements,D. and Harrison,S.C. (2005) Variable surface epitopes in the crystal structure of dengue virus type 3 envelope glycoprotein. *J. Virol.*, **79**, 1223–1231.

28. Nybakken,G.E., Oliphant,T., Johnson,S., Burke,S., Diamond,M.S. and Fremont,D.H. (2005) Structural basis of West Nile virus neutralization by a therapeutic antibody. *Nature*, **437**, 764–768.

29. Ledizet,M., Kar,K., Foellmer,H.G., Tian,W., Bushmich,S.L., Anderson,J.F., Fikrig,E. and Koski,R.A. (2005) A recombinant envelope protein vaccine against West Nile virus. *Vaccine*, **23**, 3915–3924.

30. Modis,Y., Ogata,S., Clements,D. and Harrison,S.C. (2004) Structure of the dengue virus envelope protein after membrane fusion. *Nature*, **427**, 313–319.

31. He,W., Cowin,P. and Stokes,D.L. (2003) Untangling desmosomal knots with electron tomography. *Science*, **302**, 109–113.

32. Al-Amoudi,A., Diez,D.C., Betts,M.J. and Frangakis,A.S. (2007) The molecular architecture of cadherins in native epidermal desmosomes. *Nature*, **450**, 832–837.

33. Xiong,Y. (2008) From electron microscopy to X-ray crystallography: molecular-replacement case studies. *Acta Crystallogr. D Biol. Crystallogr.*, **64**, 76–82.

34. Trapani,S., Schoehn,G., Navaza,J. and Abergel,C. (2010) Macromolecular crystal data phased by negative-stained electron-microscopy reconstructions. *Acta Cryst. D*, **66**, 514–521.

35. Baker,M.L., Jiang,W., Rixon,F.J. and Chiu,W. (2005) Common ancestry of herpesviruses and tailed DNA bacteriophages. *J. Virol.*, **79**, 14967–14970.

36. Matsumoto,A. and Ishida,H. (2009) Global conformational changes of ribosome observed by normal mode fitting for 3D CryoEM structures. *Structure*, **17**, 1605–1613.

37. Zhang,J., Baker,M.L., Schroder,G.F., Douglas,N.R., Reissmann,S., Jakana,J., Dougherty,M., Fu,C.J., Levitt,M., Ludtke,S.J. *et al.* (2010) Mechanism of folding chamber closure in a group II chaperonin. *Nature*, **463**, 379–383.

38. Wriggers,W. (2010) Using Situs for the integration of multi-resolution structures. *Biophys. Rev.*, **2**, 21–27.

39. Tama,F., Miyashita,O. and Brooks,C.L. 3rd (2004) Normal mode based flexible fitting of high-resolution structure into low-resolution experimental data from cryoEM. *J. Struct. Biol.*, **147**, 315–326.

40. Rusu,M., Birmanns,S. and Wriggers,W. (2008) Biomolecular pleiomorphism probed by spatial interpolation of coarse models. *Bioinformatics*, **24**, 2460–2466.

41. Schroder,G.F., Brunger,A.T. and Levitt,M. (2007) Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. *Structure*, **15**, 1630–1641.

42. Trabuco,L.G., Villa,E., Schreiner,E., Harrison,C.B. and Schulten,K. (2009) Molecular dynamics flexible fitting: a practical guide to combine cryo-electron microscopy and X-ray crystallography. *Methods*, **49**, 174–180.

43. Tan,R.K., Devkota,B. and Harvey,S.C. (2008) YUP.SCX: coaxing atomic models into medium resolution electron density maps. *J. Struct. Biol.*, **163**, 163–174.

44. Topf,M., Lasker,K., Webb,B., Wolfson,H., Chiu,W. and Sali,A. (2008) Protein structure fitting and refinement guided by cryoEM density. *Structure*, **16**, 295–307.

45. Baker,M.L., Ju,T. and Chiu,W. (2007) Identification of secondary structure elements in intermediate-resolution density maps. *Structure*, **15**, 7–19.

46. DiMaio,F., Tyka,M.D., Baker,M.L., Chiu,W. and Baker,D. (2009) Refinement of protein structures into low-resolution density maps using Rosetta. *J. Mol. Biol.*, **392**, 181–190.

47. Jiang,W., Baker,M., Wu,Q., Bajaj,C. and Chiu,W. (2003) Applications of bilateral denoising filter in biological electron microscopy. *J. Struct. Biol.*, **144**, 114–122.

48. Fernandez,J.J., Luque,D., Caston,J.R. and Carrascosa,J.L. (2008) Sharpening high resolution information in single particle electron cryomicroscopy. *J. Struct. Biol.*, **164**, 170–175.

49. Sousa,D. and Grigorieff,N. (2007) Ab initio resolution measurement for single particle structures. *J. Struct. Biol.*, **157**, 201–210.

50. Baker,M.L., Yu,Z., Chiu,W. and Bajaj,C. (2006) Automated segmentation of molecular subunits in electron cryomicroscopy density maps. *J. Struct. Biol.*, **156**, 432–441.

51. Pintilie,G.D., Zhang,J., Goddard,T.D., Chiu,W. and Gossard,D.C. (2010) Quantitative analysis of cryoEM density map segmentation by watershed and scale-space filtering, and fitting of structures by alignment to regions. *J. Struct. Biol.*, **170**, 427–438.

52. Goddard,T.D., Huang,C.C. and Ferrin,T.E. (2007) Visualizing density maps with UCSF Chimera. *J. Struct. Biol.*, **157**, 281–287.

53. Kaufmann,B., Chipman,P.R., Holdaway,H.A., Johnson,S., Fremont,D.H., Kuhn,R.J., Diamond,M.S. and Rossmann,M.G. (2009) Capturing a flavivirus pre-fusion intermediate. *PLoS Pathog.*, **5**, e1000672.

54. Hu,Z., Tian,X., Zhai,Y., Xu,W., Zheng,D. and Sun,F. (2010) Cryo-electron microscopy reconstructions of two types of wild rabbit hemorrahagic disease viruses characterized the structural features of Lagovirus. *Protein Cell*, **1**, 48–58.

55. Tang,J., Pan,J., Havens,W.M., Ochoa,W.F., Guu,T.S., Ghabrial,S.A., Nibert,M.L., Tao,Y.J. and Baker,T.S. (2010) Backbone trace of partitivirus capsid protein from electron cryomicroscopy and homology modeling. *Biophys. J.*, **99**, 685–694.

56. Xing,L., Li,T.C., Miyazaki,N., Simon,M.N., Wall,J.S., Moore,M., Wang,C.Y., Takeda,N., Wakita,T., Miyamura,T. *et al.* (2010) Structure of hepatitis E virion-sized particle reveals an RNA-dependent viral assembly pathway. *J. Biol. Chem*, **285**, 33175–33183.

57. Agirrezabala,X., Lei,J., Brunelle,J.L., Ortiz-Meoz,R.F., Green,R. and Frank,J. (2008) Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell*, **32**, 190–197.

58. Mayanagi,K., Kiyonari,S., Saito,M., Shirai,T., Ishino,Y. and Morikawa,K. (2009) Mechanism of replication machinery assembly as revealed by the DNA ligase-PCNA-DNA complex architecture. *Proc. Natl Acad. Sci. USA*, **106**, 4647–4652.

59. Diepholz,M., Venzke,D., Prinz,S., Batisse,C., Florchinger,B., Rossle,M., Svergun,D.I., Bottcher,B. and Fethiere,J. (2008) A different conformation for EGC stator subcomplex in solution and in the assembled yeast V-ATPase: possible implications for regulatory disassembly. *Structure*, **16**, 1789–1798.

60. Chuang,C.K., Rockel,B., Seyit,G., Walian,P.J., Schonegge,A.M., Peters,J., Zwart,P.H., Baumeister,W. and Jap,B.K. (2010) Hybrid molecular structure of the giant protease tripeptidyl peptidase II. *Nat. Struct. Mol. Biol.*, **17**, 990–996.

61. Raunser,S., Magnani,R., Huang,Z., Houtz,R.L., Trievel,R.C., Penczek,P.A. and Walz,T. (2009) Rubisco in complex with Rubisco large subunit methyltransferase. *Proc. Natl Acad. Sci. USA*, **106**, 3160–3165.

62. Rabl,J., Smith,D.M., Yu,Y., Chang,S.C., Goldberg,A.L. and Cheng,Y. (2008) Mechanism of gate opening in the 20S proteasome by the proteasomal ATPases. *Mol. Cell*, **30**, 360–368.

63. Clare,D.K. and Orlova,E.V. (2010) 4.6A CryoEM reconstruction of tobacco mosaic virus from images recorded at 300 keV on a 4k x 4k CCD camera. *J. Struct. Biol.*, **171**, 303–308.

64. Byeon,I.J., Meng,X., Jung,J., Zhao,G., Yang,R., Ahn,J., Shi,J., Concel,J., Aiken,C., Zhang,P. *et al.* (2009) Structural convergence between CryoEM and NMR reveals intersubunit interactions critical for HIV-1 capsid function. *Cell*, **139**, 780–790.

65. Kudryashev,M., Cyrklaff,M., Wallich,R., Baumeister,W. and Frischknecht,F. (2010) Distinct in situ structures of the Borrelia flagellar motor. *J. Struct. Biol.*, **169**, 54–61.

66. Huiskonen,J.T., Hepojoki,J., Laurinmaki,P., Vaheri,A., Lankinen,H., Butcher,S.J. and Grunewald,K. (2010) Electron cryotomography of Tula hantavirus suggests a unique assembly paradigm for enveloped viruses. *J. Virol.*, **84**, 4889–4897.

67. Liu,X., Zhang,Q., Murata,K., Baker,M.L., Sullivan,M.B., Fu,C., Dougherty,M.T., Schmid,M.F., Osburne,M.S., Chisholm,S.W. *et al.* (2010) Structural changes in a marine podovirus associated with release of its genome into Prochlorococcus. *Nat. Struct. Mol. Biol.*, **17**, 830–836.

68. Ranson,N.A., Farr,G.W., Roseman,A.M., Gowen,B., Fenton,W.A., Horwich,A.L. and Saibil,H.R. (2001) ATP-bound states of GroEL captured by cryo-electron microscopy. *Cell*, **107**, 869–879.

69. Liu,J., Taylor,D.W., Krementsova,E.B., Trybus,K.M. and Taylor,K.A. (2006) Three-dimensional structure of the myosin V inhibited state by cryoelectron tomography. *Nature*, **442**, 208–211.

70. Sindelar,C.V. and Downing,K.H. (2007) The beginning of kinesin's force-generating cycle visualized at 9-A resolution. *J. Cell Biol.*, **177**, 377–385.

71. Williams,K.A. (2000) Three-dimensional structure of the ion-coupled transport protein NhaA. *Nature*, **403**, 112–115.

72. Yuan,S., Yu,X., Topf,M., Ludtke,S.J., Wang,X. and Akey,C.W. (2010) Structure of an apoptosome-procaspase-9 CARD complex. *Structure*, **18**, 571–583.

73. Ludtke,S.J., Baker,M.L., Chen,D.H., Song,J.L., Chuang,D.T. and Chiu,W. (2008) De novo backbone trace of GroEL from single particle electron cryomicroscopy. *Structure*, **16**, 441–448.